

CORRECTED VERSION

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
1 August 2013 (01.08.2013)



(10) International Publication Number
WO 2013/112062 A8

(51) International Patent Classification:
H04L 12/58 (2006.01)

(21) International Application Number:
PCT/RO2012/000023

(22) International Filing Date:
5 September 2012 (05.09.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/358,358 25 January 2012 (25.01.2012) US

(71) Applicant (for all designated States except US): **BITDEFENDER IPR MANAGEMENT LTD** [—/CY]; Kreontos 12, PC 1076, Nicosia (CY).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **DICHIU, Daniel** [RO/RO]; Str. Piata Amzei nr. 10-22, sc. D, et 7, ap. 73, sector 1, Bucuresti (RO). **LUPESCU Z. Lucian** [RO/RO]; Str. Nița Elineanu Nr. 60A, Sector 3, București (RO).

(74) Agent: **TULUCA, Doina**; Bd. Lacul Tei nr. 56, bl. 19, sc.B, ap. 52, sector 2, RO-020392 Bucuresti (RO).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

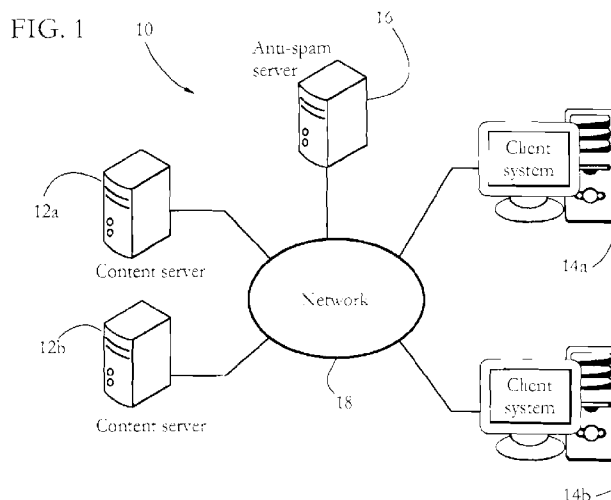
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

- with international search report (Art. 21(3))

[Continued on next page]

(54) Title: SYSTEMS AND METHODS FOR SPAM DETECTION USING CHARACTER HISTOGRAMS



(57) Abstract: Described spam detection techniques including string identification, pre-filtering, and character histogram and timestamp comparison steps facilitate accurate, computationally-efficient detection of rapidly-changing spam arriving in short-lasting waves. In some embodiments, a computer system extracts a target character string from an electronic communication such as a blog comment, transmits it to an anti-spam server, and receives an indicator of whether the respective electronic communication is spam or non-spam from the anti-spam server. The anti-spam server determines whether the electronic communication is spam or non-spam according to certain features of the character histogram of the target string. Some embodiments also perform an unsupervised clustering of incoming target strings into clusters, wherein all members of a cluster have similar character histograms.



(48) Date of publication of this corrected version:

17 July 2014

(15) Information about Correction:

see Notice of 17 July 2014